# Pathomic prediction of CCL2 expression and its prognostic significance in breast cancer

Ze-Juan Yang[1]*, Lan Luo[2], Hui-Feng Wang[3]*

[1]Department of Breast Surgery, The Affiliated TCM Hospital of Guangzhou Medical University; Clinical College of TCM and Western Medicine, Guangzhou Medical University, Guangzhou, Guangdong Province, China
[2]The Affiliated Guangzhou Hospital of TCM of Guangzhou University of Chinese Medicine, Guangzhou, Guangdong Province, China
[3]Department of Human Anatomy, School of Basic Medical Sciences, Guangxi Medical University, Nanning, Guangxi Province, China

**Submitted:** 24 September 2026; **Accepted:** 30 September 2026
**Online publication:** 30 September 2026

Arch Med Sci 2026; 22 (5): 3101–3105
DOI: https://doi.org/10.5114/aoms/242431
Copyright © 2026 Termedia & Banach

***Corresponding authors:**
Ze-Juan Yang Prof.
Department of Breast Surgery
The Affiliated TCM Hospital of Guangzhou Medical University
Clinical College of TCM and Western Medicine
Guangzhou Medical University
Guangzhou 510130
Guangdong Province
China. E-mail: dryang606@163.com

Hui-Feng Wang
Department of Human Anatomy
School of Basic Medical Sciences
Guangxi Medical University
Nanning 530021
Guangxi Province, China
E-mail: wanghuifeng@gxmu.com

Breast cancer is the most common malignant tumour and the second leading cause of cancer-related death among women worldwide [1]. Surgery, radiotherapy, chemotherapy, targeted therapy and immunotherapy are the mainstay treatments [2], yet 5-year survival ranges from 85–90% in developed countries to 54–74% in developing countries [3]. Mammography, ultrasonography and magnetic resonance imaging are widely used, but false positives and inter-observer variability limit accuracy [4]. New prognostic markers are therefore needed to stratify patients and guide personalised treatment.

C-C motif chemokine ligand 2 (CCL2), also known as monocyte chemoattractant protein-1, belongs to the CC chemokine family and signals through CCR2 and CCR4 [5]. CCL2 is a key regulator of the chemokine-driven tumour microenvironment: it enhances cancer-cell survival by modulating apoptosis, necrosis and autophagy, and recruits immune cells in a context-dependent manner [6]. CCL2 may also be protective by recruiting anti-tumour T-cell subsets into the tumour microenvironment [7], suggesting dual, context-dependent effects on tumours [8].

At present, CCL2 expression is assessed mainly by blood cytokine measurement or by mRNA or protein assays on fresh or paraffin-embedded tissue; these methods are invasive, operator- and antibody-dependent, and costly [6, 9]. Whether pathomics can evaluate CCL2 expression non-invasively remains unclear. We therefore analysed the association between CCL2 expression and prognosis, explored potential mechanisms, and developed a pathomics model as a practical non-invasive imaging biomarker.

**Methods.** ***Patient cohorts.*** We retrospectively analysed transcriptomic, clinical, follow-up and histopathological data from The Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA) cohort. After excluding patients with metastatic disease, survival < 30 days or missing survival data, incomplete clinical data, male sex, other tumour sites, missing RNA-seq data or poor-quality images, 791 patients were available for imaging-genomics analysis. Using a data-driven cutoff of 3.4491, determined by the minimum *p*-value approach (scanning all candidate cutoffs to maximise the log-rank statistic), patients were divided into high-CCL2 ($n$ = 375) and low-CCL2 ($n$ = 416) groups and randomly split into training

Creative Commons licenses: This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY -NC -SA 4.0). License (http://creativecommons.org/licenses/by-nc-sa/4.0/).

($n$ = 633) and validation ($n$ = 158) sets at an 8 : 2 ratio. Level-3 HTSeq-FPKM RNA-seq data and 113 GTEx normal-tissue samples were processed with the Toil pipeline and log2-transformed.

***Survival analysis.*** Kaplan-Meier curves and the log-rank test compared survival between groups. Univariate and multivariate Cox proportional hazards models assessed factors associated with overall survival (OS). Subgroup analyses using univariate Cox regression evaluated the effect of CCL2 within covariate strata, and likelihood-ratio tests assessed interactions with other covariates.

***Pathomics workflow.*** The pipeline comprised pre-processing, segmentation, feature extraction, feature selection and evaluation (Figure 1). Haematoxylin and eosin (H&E)-stained whole-slide images (SVS format) were downloaded from TCGA at 20× or 40× magnification [10]. The OTSU algorithm separated tissue from background; 40× images were divided into 1024 × 1024-pixel sub-images and 20x images into 512 × 512-pixel sub-images upsampled to 1024 × 1024 pixels. After excluding sub-images with contamination, blurring or > 50% blank areas, ten random sub-images per slide were selected to reduce redundancy while capturing intratumoural heterogeneity [11, 12]. PyRadiomics extracted 93 original features, and wavelet transformation yielded 465 in total; mRMR and recursive feature elimination (RFE) selected seven generalisable features for a Gradient Boosting Machine (GBM) model predicting CCL2 expression [13]. Performance was evaluated by ROC curves, calibration curves, the Hosmer-Lemeshow test, Brier score and decision curve analysis (DCA). With a pathomics score (PS) cutoff of 0.5127, determined by the same approach, patients were classified into low-PS ($n$ = 444) and high-PS ($n$ = 347) groups.

***Mechanistic analysis and statistics.*** Gene set variation analysis (GSVA) calculated enrichment scores for KEGG (186 pathways) and Hallmark (50 pathways) gene sets; differential analysis between PS groups used "limma". Spearman correlation assessed the relationship between PS and chemokine gene expression. CIBERSORTx estimated the fractions of 22 immune cell subtypes, and tumour mutational burden (TMB) was calculated with Maftools. The Wilcoxon rank-sum and $\chi^2$ tests were used as appropriate; analyses were performed in R 3.4.4, and $p < 0.05$ was considered significant. All tests were two-sided. The pathway, chemokine and immune infiltration analyses were exploratory, with no adjustment for multiple testing, and are therefore hypothesis-generating.

**Results.** ***CCL2 expression and clinicopathological features.*** Among 791 patients, ER status, PR status, HER2 status, T stage and chemotherapy differed significantly between CCL2 groups (all $p < 0.05$). CCL2 expression was higher in tumours than in normal tissue ($p = 0.028$).

***CCL2 expression and survival.*** The high-CCL2 group showed a non-significant trend towards better OS than the low-CCL2 group (median OS 215.2 vs. 130.9 months; $p = 0.072$). In univariate analysis, CCL2 was not significantly protective (HR = 0.705, 95% CI: 0.480–1.034, $p = 0.073$), whereas ER status, PR status, radiotherapy and chemotherapy were protective. After multivariate adjustment, CCL2 remained independently associated with better OS (HR = 0.578, 95% CI: 0.375–0.889, $p = 0.013$); PR status, radiotherapy and chemotherapy were also protective, while age, N stage and margin status were risk factors. Subgroup analyses showed protective associations for ER status, HER2 status and T2 stage (all $p < 0.05$, except HER2: $p = 0.051$).

***Pathomics model for CCL2 expression.*** Seven features (four original, three wavelet) with good generalisability were selected (Figures 1 A, B); the most discriminative were original_gldm_LargeDependenceLowGrayLevelEmphasis, wavelet_HL_firstorder_Skewness and wavelet_LH_glcm_Correlation. The model achieved AUCs of 0.825 (training) and 0.730 (validation), with calibration curves showing good agreement between predicted and observed values and DCA indicating clinical utility (Figure 1 D). PS was higher in the high-CCL2 group in both sets ($p < 0.0001$).

***PS and survival.*** The high-PS group had better OS than the low-PS group (median OS 215.2 vs. 129.1 months; $p = 0.006$) (Figure 1 C). PS was protective in univariate analysis (HR = 0.583, 95% CI: 0.394–0.862, $p = 0.007$) and remained an independent protective factor after multivariate adjustment (HR = 0.498, 95% CI: 0.324–0.765, $p = 0.001$); PR status, radiotherapy and chemotherapy were also protective, whereas age, N stage and margin status were risk factors. No significant interaction between PS and the tested covariates was observed.

***PS, pathway enrichment and immune infiltration.*** GSVA showed that the high-PS group was positively enriched for VEGF signalling (KEGG), and for P53 pathway, apoptosis, HEDGEHOG and NOTCH signalling (Hallmark), and negatively enriched for cell cycle, DNA replication, MYC targets v1/v2, G2M checkpoint and E2F targets (Figure 1 E). PS correlated positively with CCL4, CCL8 and CCL13 (all $p < 0.001$) and negatively with SEMA3B, SEMA3F and SEMA4G ($p < 0.05$). The high-PS group also showed greater infiltration of naïve B cells, CD8+ T cells, gamma delta T cells and M2 macrophages, and lower infiltration of resting memory CD4+ T cells and M1 macrophages (all $p < 0.05$) (Figure 1 E). TMB was higher in the high-PS group ($p < 0.01$) (Figure 1 E).

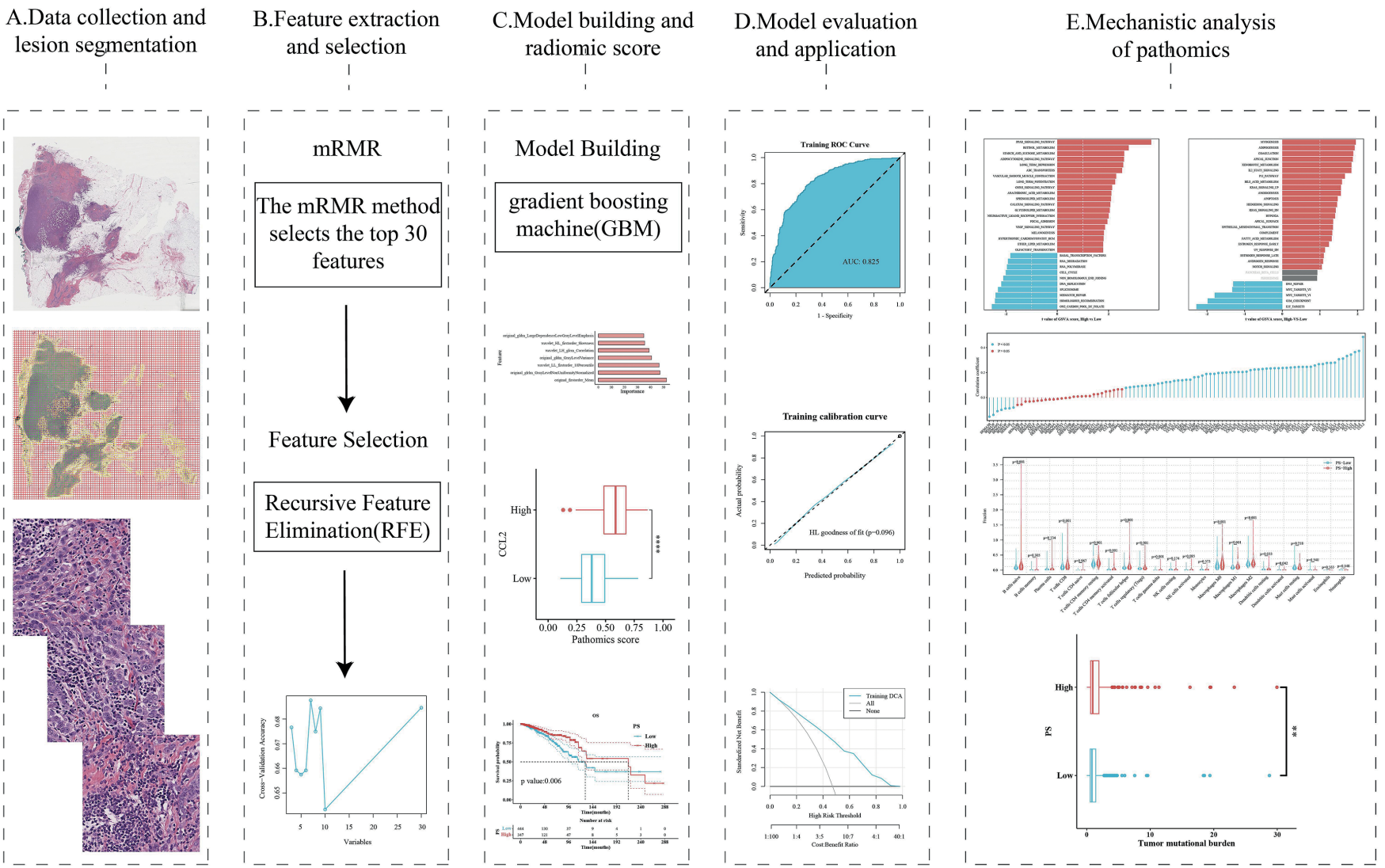


**Figure 1.** Pathomics model for CCL2 expression and prognosis in breast cancer. **A** – Data collection and lesion segmentation: haematoxylin and eosin-stained whole-slide images were segmented and divided into sub-images. **B** – Feature extraction and selection using minimum redundancy maximum relevance (mRMR) and recursive feature elimination (RFE). **C** – Model building with a Gradient Boosting Machine: importance of the seven selected features, pathomics score (PS) distribution between high- and low-CCL2 groups, and Kaplan-Meier curves for overall survival according to the PS group. **D** – Model evaluation: receiver operating characteristic curves in the training (AUC = 0.825) and validation (AUC = 0.730) sets, calibration curves, Hosmer-Lemeshow test and decision curve analysis. **E** – Mechanistic analysis: gene set variation analysis of KEGG and Hallmark gene sets, correlations between PS and chemokines, immune cell infiltration, and tumour mutational burden between high- and low-PS groups

*AUC – area under the curve, CCL2 – C-C motif chemokine ligand 2, PS – pathomics score.*

**Discussion.** CCL2 attracts monocytes and other immune cells to the tumour site and has attracted considerable attention in breast cancer research [14]; its role remains contradictory, with both protective and harmful effects reported. In our study, CCL2 was an independent protective factor for OS (HR = 0.578, 95% CI: 0.375–0.889, $p$ = 0.013), with no significant interaction between CCL2 and ER, PR or HER2 status. This counter-intuitive finding reflects the data source: TCGA provides bulk rather than single-cell transcriptomes, and CCL2 in bulk tissue derives largely from stromal and immune cells rather than malignant cells. Higher CCL2 levels may therefore mark a more extensively immune-infiltrated microenvironment, generally associated with favourable outcomes, rather than a tumour-cell-intrinsic effect; the greater immune infiltration in the high-PS group supports this view.

Conversely, CCL2 can promote progression by recruiting regulatory T cells, inflammatory monocytes and macrophages and shaping an immunosuppressive microenvironment [15], and it predicts poor outcomes in other tumour types [9]. Distinguishing stromal from malignant cell-of-origin may reconcile these opposing observations and explain why CCL2 has not entered routine prognostic use.

Unlike studies restricted to a single molecular subtype, we predicted CCL2 expression non-invasively from routine H&E slides, with good calibration and clinical utility. The high-PS group was enriched for VEGF, P53, apoptosis and NOTCH signalling, and PS correlated with CCL4, CCL8 and CCL13, chemokines sharing the CCR2 axis with CCL2 that may cooperate in remodelling the tumour microenvironment [5, 7], together with greater immune infiltration and higher TMB. These findings suggest that CCL2 may influence breast cancer growth partly through chemokine- and immune-mediated regulation of apoptosis, but remain hypothesis-generating and require experimental confirmation.

This study has limitations. It was retrospective and based on a single public database, and multi-centre external validation is needed; notably, the pathomics model could not be validated externally because no independent publicly available breast cancer cohort provides matched H&E whole-slide images, transcriptomic data and survival follow-up. The minimum $p$-value approach selects the most discriminative cut-point and may yield optimistic $p$-values and overestimated effect sizes, so the borderline univariate association ($p$ = 0.073) warrants caution and the conclusion of independent prognostic value rests on the multivariate model. Image-derived features may also vary with scanner, magnification and staining protocol.

In conclusion, the pathomics model provided a practical, non-invasive means of estimating CCL2 expression from routine H&E slides, and both CCL2 expression and PS were independently associated with better OS in breast cancer. External validation is required before clinical application.

## Funding

Wu Jieping Medical Foundation (Grant No. 320.6750.2024-03-21) Guangzhou Municipal Science and Technology Bureau (Grant No. 2025A03J3160) Scientific Research Strengthening Foundation of Guangzhou Medical University (Grant No. 2025SRP030) Traditional Chinese Medicine Organoid Laboratory Project of the Affiliated Traditional Chinese Medicine Hospital of Guangzhou Medical University (Grant No. LQG2026004).

## Ethical approval

Not applicable.

## Conflict of interest

The authors declare no conflict of interest.

## References

1. Tang DD, Ye ZJ, Liu WW, et al. Survival feature and trend of female breast cancer: a comprehensive review of survival analysis from cancer registration data. Breast 2025; 79: 103862.
2. Wu S, Chen P, Zhou W, et al. Precision targeted therapy and immunotherapy in breast cancer: advances and clinical frontiers. Chin Med J 2026. doi:10.1097/CM9.0000000000004168.
3. Fink H, Soerjomataram I, Bardot A, et al. International comparisons of survival after recurrent metastatic breast cancer in four countries: a population-based study. Breast 2026; 87: 104778.
4. Uwimana A, Gnecco G, Riccaboni M. Artificial intelligence for breast cancer detection and its health technology assessment: a scoping review. Comput Biol Med 2025; 184: 109391.
5. Lim SY, Yuzhalin AE, Gordon-Weeks AN, Muschel RJ. Targeting the CCL2-CCR2 signalling axis in cancer metastasis. Oncotarget 2016; 7: 28697-710.
6. Rogic A, Pant I, Grumolato L, et al. High endogenous CCL2 expression promotes the aggressive phenotype of human inflammatory breast cancer. Nat Commun 2021; 12: 6889.
7. Lanca T, Costa MF, Goncalves-Sousa N, et al. Protective role of the inflammatory CCR2/CCL2 chemokine pathway through recruitment of type 1 cytotoxic gammadelta T lymphocytes to tumour beds. J Immunol 2013; 190: 6673-80.
8. Lavender N, Yang J, Chen SC, et al. The Yin/Yang of CCL2: a minor role in neutrophil anti-tumour activity in vitro but a major role on the outgrowth of metastatic breast cancer lesions in the lung in vivo. BMC Cancer 2017; 17: 88.
9. Lu X, Qian CN, Mu YG, et al. Serum CCL2 and serum TNF-alpha – two new biomarkers predict bone invasion, post-treatment distant metastasis and poor overall sur-

vival in nasopharyngeal carcinoma. Eur J Cancer 2011; 47: 339-46.
10. Nishio M, Nishio M, Jimbo N, Nakane K. Homology-based image processing for automatic classification of histopathological images of lung tissue. Cancers (Basel) 2021; 13: 1192.
11. Chen L, Zeng H, Zhang M, Luo Y, Ma X. Histopathological image and gene expression pattern analysis for predicting molecular features and prognosis of head and neck squamous cell carcinoma. Cancer Med 2021; 10: 4615-28.
12. Zeng H, Chen L, Zhang M, Luo Y, Ma X. Integration of histopathological images and multi-dimensional omics analyses predicts molecular features and prognosis in high-grade serous ovarian cancer. Gynecol Oncol 2021; 163: 171-80.
13. Saednia K, Lagree A, Alera MA, et al. Quantitative digital histopathology and machine learning to predict pathological complete response to chemotherapy in breast cancer patients using pre-treatment tumour biopsies. Sci Rep 2022; 12: 9690.
14. Yoshimura T, Li C, Wang Y, Matsukawa A. The chemokine monocyte chemoattractant protein-1/CCL2 is a promoter of breast cancer metastasis. Cell Mol Immunol 2023; 20: 714-38.
15. Archer M, Bernhardt SM, Hodson LJ, et al. CCL2-mediated stromal interactions drive macrophage polarisation to increase breast tumorigenesis. Int J Mol Sci 2023; 24: 7385.